National Climatic Data Center

DATA DOCUMENTATION

FOR

DATA SET 9641K (DSI-9641K)

NORMALS OF SNOWFALL AND SNOW DEPTH, 1961-1990

December 23, 2002

National Climatic Data Center
151 Patton Ave.
Asheville, NC 28801-5001 USA

:
:

Table of Contents

1. Abstract: The purpose of this project was to prepare official snowfall and snow depth normals for the period 1961-1990 for weather stations across the United States and its territories. U.S. climate normals are periodically computed by the NCDC according to guidelines established by the World Meteorological Organization (WMO). QC'd data from the Production of Snowfall and Snow Depth Climatologies for NWS Cooperative Observer Sites project were used in this normals project for Cooperative (COOP) stations, while normals previously computed from another data base were used for First Order/airport (WBAN) stations.

There are 3295 stations in this data set, of which 258 are First Order/airport stations and 3037 are volunteer Cooperative stations. All of the stations are identified by a Cooperative Identification Number. The snow normals were computed for stations that had at least 15 years of non-missing data for total snowfall, daily snowfall and snow cover, and number of days with snowfall and snow cover during the normals period.

The ValHiDD QC methodology may have compromised, in some cases, the edited DSI-3200 snow data values for WBAN stations (see Quality Statement). Fortunately, snowfall and snow depth normals for WBAN stations had already been computed using data from NCDC's DSI-3210 First Order Summary of the Day data base, which had not been subjected to ValHiDD processing. These WBAN snow normals were extracted from the archive and inserted into the final product for this normals project.

COMPUTATIONAL METHODOLOGY

Data from 1961-1990 were analyzed. A suite of statistics (mean, median, first and third quartiles, and extremes [both amounts and dates of occurrence]) was generated for several climatic parameters. The specific statistics that were computed vary with parameter, but the number of years with non-missing data (NYRS) was computed for each parameter. The NYRS information is very important for any inter-station comparisons the user may wish to make.

Sequential year-month values were generated for each year and month from January 1961 through December 1990. The monthly normals were computed from these year-month sequential values. WMO guidance was followed in the computation of the normals. The sequential year-month data were not adjusted for inhomogeneities.

CLIMATIC PARAMETERS

The climatic parameters include the following:

(1) number of days with daily snowfall amount equal to zero or a trace;

(2) number of days with daily snowfall amount greater than or equal to several thresholds (0.1, 1.0, 2.0, 5.0, and 10.0 inches);

(3) monthly and annual total snowfall amount;

(4) daily snowfall amount, both with all days examined (whether they had snowfall or not) and only days having snowfall examined;

(5) number of days with snow depth equal to zero or a trace;

:

(6) number of days with snow depth greater than or equal to several thresholds (1.0, 2.0, 5.0, and 10.0 inches); and

(7) daily snow depth amount, both with all days examined (whether they had snow cover or not) and with only days with snow cover examined.

MONTHLY/ANNUAL CLIMATOLOGY COMPUTATIONAL CONSIDERATIONS

The date (year, month, and/or day) of an extreme is the date of the most recent occurrence.

Monthly statistics (for January through December) were computed based on the days in the month under consideration. The annual value (provided for just the mean statistic) was computed by summing the monthly normal values. The annual value may not exactly equal the sum of the monthly values due to rounding errors. For the daily snowfall (snow depth) amount, where only days having snowfall (snow depth) were examined (groups 4 and 7 in above section), if no snow occurred, then there was no data from which to compute a value. Consequently, the NYRS statistics (for the COOP stations) refers to the number of years with non-zero data, which is somewhat broader than the number of years with non-missing data.

The impact of missing data varies, depending on the element, the statistic computed, and the type of station.

THE EFFECT OF MISSING DATA: COOP STATIONS

Total snowfall had no tolerance for missing data. If even one day was missing in a month, the total snowfall could not be computed for that year=s month. Consequently, the number of years with non-missing data will vary with month.

The median daily value for a month had no tolerance for missing data. All days in a month had to have data in order for a median daily value to be computed for that year-month.

The number of days with snowfall or snow depth parameters had no tolerance for missing data. Data for leap days (February 29) were included in the analysis. Due to this fact and due to rounding error, the sum of the values for the equal zero, equal trace, and greater than or equal to 0.1 inch (for snowfall, 1.0 inch for snow depth) thresholds may not exactly equal the maximum possible number of days in the month.

The daily extreme and date of occurrence parameters had a greater tolerance for missing data. Data were analyzed even if a month had up to 5 days missing. This could result in apparent discrepancies between these and other parameters.

THE EFFECT OF MISSING DATA: WBAN STATIONS

For any given year and month in the sequential data, the monthly value was not computed from the daily values (the month was considered missing) if (1) five or more consecutive daily values were missing or (2) eleven or more daily values in total in the month were missing.

In the "number of days with" computations, if a month had a few missing days but still had sufficient days to be included in the analysis, then the number of days tallies were pro-rated by a factor consisting of (pos/obs), where pos

:

= the number of days possible in the month and obs = the number of days in the month with non-missing data. Data for February 29 (leap days) was included in the analysis, but the results were pro-rated as above to conform to a 28-day February. The pro-rating feature, and possible rounding errors, may result in the following apparent inconsistencies: (1) the sum of the monthly values may not exactly equal the annual value, and (2) the cross element sums (e.g., number of days with snowfall [or snow depth] = zero, plus number of days = trace, plus number of days >= 0.1 inch [for snowfall, 1.0 inch for snow depth]) may not exactly equal the maximum possible monthly (28.0, 30.0, or 31.0) or annual (365.0) value. In all cases, the apparent inconsistencies are minor.

The source data for this data set originated from the Summary of the Day/Surface Land Daily Cooperative (DSI-3200) data set and the Summary of the Day/First Order (DSI-3210) data set.

The snow normals for the WBAN stations were originally prepared for the DSI-9641 data set: U.S. Stations 1961-1990 Monthly Normals for the Atypical Climate Elements [Wind, Pressure, Humidity, Snow, Cloud Cover, Sunshine, Days with Weather].

2. **Element Names and Definitions**:

This data set consists of two files on one magnetic tape. The first file contains station metadata. The second file contains 1961-1990 monthly and annual normals for snowfall and snow depth.

There are 3295 stations in this data set, of which 258 are First Order/airport stations and 3037 are volunteer Cooperative stations. All of the stations are identified by a Cooperative identification number.

The data in each of the two files have the following specifications:

Record Length	: Fixed 92 characters
Blocked	: 4600 characters
Media	: ASCII 18-Track IBM-Type 3480 cartridge
Parity	: Odd
Label	: ANSI Standard Labeled

METADATA FILE

The stations in the 1961-1990 snow normals data set, and the normals statistics that were computed, are a mixture from two different source data sets: Cooperative (COOP) stations, derived using data from NCDC's DSI-3200 Cooperative Summary of the Day data base; and First Order/airport (WBAN) stations, derived using data from NCDC's DSI-3210 First Order Summary of the Day data base. All of the stations have an assigned COOP number, while only the First Order/airport stations have an assigned WBAN number. For this snow normals project, the Cooperative Station History File was the source for all of the metadata information (except station name and WBAN number for the WBAN stations, which were taken from the DSI-9641 data set, AU.S. Stations 1961-1990 Monthly Normals for the Atypical Climate Elements [Wind, Pressure, Humidity, Snow, Cloud Cover, Sunshine, Days with Weather].

The station metadata are archived by ascending station COOP number sort. Each station metadata record has the following record format:

:
:

POSITION	WIDTH	TYPE	CONTENTS
001-006	6	numeric	COOP station identification number
007-007	1	alpha	not used
008-012	5	numeric	WBAN station identification number blank = WBAN # not assigned
013-013	1	alpha	not used
014-014	1	numeric	Data Source Code 0 = DSI-3200 1 = DSI-3210
015-015	1	alpha	not used
016-045	30	alpha-num	station name
046-046	1	alpha	not used
047-048	2	alpha	two-letter state abbreviation
049-049	1	alpha	not used
050-054	5	alpha-num	latitude (ddmmh, where dd=degrees, mm=minutes, h=hemisphere [N for North, S for South])
055-055	1	alpha	not used
056-061	6	alpha-num	longitude (dddmmh, where ddd=degrees, mm=minutes, h=hemisphere [W for West, E for East])
062-062	1	alpha	not used
063-067	5	numeric	elevation, to whole feet -999 = not available
068-092	25	alpha	not used

DATA FILE

The 1961-1990 snowfall and snow depth normals output data are archived by ascending station number sort, then by parameter (element code, threshold code, statistic code, date indicator code). Each data record consists of station identification information, data period, parameter information, twelve monthly data values, and one annual data value.

A four-tiered parameter coding system was developed to accommodate the large variety of parameters and statistics that were computed. The parameter coding system consists of a climatic element code, a statistic code, a date indicator code (which identifies the values as data values, dates, or years), and a threshold code.

The data consist of decimal values, integer values, and special codes. Snowfall and snow depth amount are in inches. The special codes are defined as follows:

CODE VALUE	DEFINITION
-99.9	code for missing value due to data not available, insufficient data to compute a value, or statistic not applicable for this month or season (annual)
-99	code for missing value or statistic not applicable for this month or season
-8.8	observed trace value
-7.7	code for computed value > 0.0 but < 0.5 (for snow depth) or 0.05 (for snowfall and days with), based on measurable observations

:
:

Each data record has the following record format:

POSITION	WIDTH	TYPE	CONTENTS
001-006	6	numeric	Cooperative Network station number
007-010	4	numeric	first year of input data analyzed
011-014	4	numeric	last year of input data analyzed
015-016	2	numeric	climatic element code (see Table 1)
017-018	2	alpha-num	statistic code (see Table 2)
019-019	1	alpha	date indicator code (see Table 3)
020-023	4	alpha-num	threshold code (see Table 4)
024-025	2	alpha	not used
026-030	5	numeric	January value
031-035	5	numeric	February value
036-040	5	numeric	March value
041-045	5	numeric	April value
046-050	5	numeric	May value
051-055	5	numeric	June value
056-060	5	numeric	July value
061-065	5	numeric	August value
066-070	5	numeric	September value
071-075	5	numeric	October value
076-080	5	numeric	November value
081-085	5	numeric	December value
086-091	6	numeric	Annual value
092-092	1	alpha	not used

3. **Start Date**: 19610101

4. **Stop Date**: 19901231

5. **Coverage**: North America and parts of the Pacific Ocean

a. Southernmost Latitude:	24N	Data sets other than Normals
	15S	Normals data set
b. Northernmost Latitude:	75N	All data sets including Normals
c. Westernmost Longitude:	171W	Data sets other than Normals
	130E	Normals data set
d. Easternmost Longitude:	67W	Data sets other than Normals
	65W	Normals data set

6. **How to Order Data**:

Ask NCDC's Climate Services about the cost of obtaining this data set.
Phone: 828-271-4800
FAX: 828-271-4876
E-mail: NCDC.Orders@noaa.gov

7. **Archiving Data Center**:

National Climatic Data Center
Federal Building
151 Patton Avenue
Asheville, NC 28801-5001
Phone: (828) 271-4800.

:
:

8. Technical Contact:

National Climatic Data Center
Federal Building
151 Patton Avenue
Asheville, NC 28801-5001
Phone: (828) 271-4800.

9. Known Uncorrected Problems: The properties of snow make it difficult to accurately and consistently measure snowfall. Snow often melts as it lands or as it lies on the ground, snow settles as it lies on the ground, and snow is easily blown and redistributed. These properties can be affected by location, time of day the observations are taken, and how often they are measured (Doesken and Judson, 1997). For these reasons, it is important for observers to adhere to a standard methodology (National Weather Service, 1972) for observing and reporting snow. Unfortunately, stations change location, observers, and sometimes observation time. Such changes introduce inhomogeneities into the snow record. No acceptable adjustment algorithms exist to statistically adjust daily snow data for inhomogeneities.

10. Quality Statement: Different quality control (QC) methodologies were employed for the two types (WBAN and COOP) of stations.

QUALITY CONTROL OF THE WBAN DATA

The WBAN stations' quality assurance process consisted of both manual and automated steps and was applied to the DSI-3210 data. The data were displayed by means of an array that showed the frequency of occurrences along the entire range of observed values. This array was manually examined, with outliers being identified and flagged. If an outlier was determined to be questionable, then it was deleted from the analysis. A general knowledge of potential extremes as well as the individual station's climatology provided guidance in this process. A Hi-Lo file (a list of threshold values at both the upper and lower extremes revealed by the visual array process) was used in the automated step to filter the daily data, with values beyond the Hi-Lo thresholds being deleted.

QUALITY CONTROL OF THE COOP DATA

Three levels of quality control were employed in order to obtain the best snow data possible. The first level involved using the ValHiDD edited DSI-3200 values (Reek, *et al.*, 1992). The second level employed a number of additional internal consistency checks wherein the daily snowfall was compared to the corresponding maximum and minimum temperature, precipitation, and snow depth. The third level was an extremes check. Snowfall and snow depth values that failed the checks were corrected (where possible) or set to missing.

First Level QC: ValHiDD

During the 1990's, an automated quality control system called ValHiDD (Validation of Historical Daily Data) was applied to the entire DSI-3200 database to remove gross errors in daily maximum and minimum temperature, precipitation, snowfall, and snow depth. ValHiDD is a rules-based method for detecting and correcting discrepancies (due to digitizing errors and observer errors) in the DSI-3200 database. The checks employed by ValHiDD include a limits check, internal consistency checks, flatliner temperature check, precipitation/snowfall/snow depth (PSFSD) relationship check, temperature spike check, multiple rule-group failures check, and failed fix check (Reek, *et al.*, 1992).

Although the number of discrepancies uncovered and resolved by ValHiDD was small compared to the total number of data values examined, their removal/correction was important and ValHiDD significantly contributed to the improvement of the overall DSI-3200 database. However, two factors relevant to this project should be noted:

(1) In some PSFSD cases, ValHiDD could not identify which element should be corrected, so the values were flagged as suspect and not altered.

(2) The PSFSD relationship check assumed that all three elements were observed at the same hour. For most volunteer COOP observers this assumption holds. However, for airport stations, this is not the case: snow depth is observed at 7 a.m. local time, while daily snowfall and precipitation amount are reported as of midnight. This airport station observation time discrepancy complicated the PSFSD relationship check.

Second Level QC: Internal Consistency Checks

This level of QC included temporal checks (comparing today=s snow depth values to yesterday=s values) and additional inter-element checks beyond those performed by ValHiDD. Snowfall and snow depth values that failed the internal consistency checks were corrected (where possible) or set to missing. Temperature or precipitation values were not examined for accuracy at this level.

The second level QC included the following checks. The following abbreviations are used here: TMIN = minimum temperature (deg. F), TMAX = maximum temperature (deg. F), P = precipitation (inches), SF = snowfall (inches), and SD = snow depth (inches).

(1) Factor of 10 error for SF: if P >= 0.01 and SF >= 1.0 and the ratio, SF/P, was greater than 80.0, then the SF value was corrected by dividing by 10. The corrected SF value was similarly checked and set to missing if the new SF/P ratio was greater than 50.

(2) Hail check: non-zero SF values were set to zero if TMIN >= 40.

An alternative check was used for those cases where the minimum temperature was missing (stations measuring both temperature and precipitation where the day=s TMIN was missing, and stations which measured only precipitation). This alternative involved examining the day=s climatological median extreme minimum temperature (CMEMT) for the state. The CMEMT was computed for each of the 365 days of the year (the value for February 28 was used for February 29 leap days) for each state from the daily extreme minimum temperature values for each station in the state, from the period 1961-1990. Non-zero SF values were set to zero if CMEMT > 25.

(3) Non-zero SF values were set to missing if:
- (I) SF > 0.4 but P = 0; or
- (II) today’s P is missing.

(4) Factor of 10 error for SD (where previous day’s SD = zero or trace): SD

:

was compared to SF and corrected if it was identified as being off by a factor of 10. If the SD was greater than ten times SF, the SD was set to missing. (There were a few cases where the observer inconsistently recorded SD off by a factor of ten for a string of years. This check was used to identify the beginning and ending years of such periods, so the station's data could be later examined manually. If the error was not consistent, the snow depth from this string of years was subsequently deleted from the analysis.)

(5) Second check for factor of 10 error for SD: if the difference between today's SD and yesterday's SD was greater than today's SF (plus an adjustment factor due to difference in units resolution), today's SD was divided by 10. The corrected SD value was similarly checked and set to missing if the difference in SD was still greater than today's SF.

(6) Zero SD values were set to missing if: yesterday's SD > 7 and today's SF > 2.0.

(7) Non-zero SD values were set to missing if:
- (I) today's SD > yesterday's SD with today's SF = 0; or
- (II) today's SF is missing; or
- (III) yesterday's SD missing and today's SD > (today's SF + SD of last day with non-missing SD).

(8) Today=s SD was set to missing if today's P < 0.05 and:
- (I) yesterday's SD >= (4 + today's SD), and today's TMAX < 40; or
- (II) yesterday's SD >= (7 + today's SD), and today's TMAX < 45; or
- (III) yesterday's SD >= (10 + today's SD), and today's TMAX > 44; or
- (IV) yesterday's SD >= (7 + today's SD), and today's TMAX missing.

As in the ValHiDD discussion (see section 16.2.1), it is critical to these automated tests that the temperature, precipitation, snowfall, and snow depth observations be taken at the same hour. This is the case for most volunteer COOP observers. However, National Weather Service (NWS) and Federal Aviation Administration (FAA) airport stations observe snow depth at 7 a.m. local time, while the remaining elements are reported as of midnight. This airport station observation time discrepancy impacts checks (4)-(7) above, and can result in valid snow depth values being flagged as erroneous and being changed.

Extremes Checks

The daily snowfall values were compared, on a state-by-state basis, to known statewide 24-hour snowfall extremes. The known extremes published by weather historian David Ludlum (Ludlum, 1982) were, for most states, multiplied by an acceptability factor (1.4) in order to account for new daily extremes that may have been set since his book was published, and to account for the difference in time frame (a moving 24-hour time frame versus daily values taken at a fixed ob time). Special subjective estimates were used for Colorado, Florida, and New York.

These adjusted statewide extremes were used in the snowfall extremes check. If a station's daily snowfall value exceeded the corresponding statewide extreme, the value was set to missing.

Snow depth varies widely in states with mountain topography. For example, the extremes for coastal stations in southern California would be considerably lower than the extremes for stations in the Sierra Nevada range. This made it difficult to establish an appropriate statewide snow depth extreme, so a standard snow depth extreme of 2000 inches was used for all stations. If a station's snow depth value exceeded 2000 inches, the value was set to missing.

11. Essential Companion Datasets: None.

12. References:

Doesken, N.J. and A. Judson, 1997: ***A Guide to the Science, Climatology, and Measurement of Snow in the United States***, Second Edition, Colorado State University Department of Atmospheric Science: Fort Collins.

Ludlum, D.M., 1982: ***The American Weather Book***, Houghton Mifflin Co.: Boston.

National Weather Service, 1972: ***National Weather Service Observing Handbook No. 2: Substation Observations***, First Edition, Revised December 1972 (Supersedes Circular B), U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, Silver Spring, MD.

Reek, T., S.R. Doty, and T.W. Owen, 1992: AA deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network.@ ***Bulletin of the American Meteorological Society***, vol. 73, pp. 753-762.

:

Appendix: Data and Metadata Code Tables

Table 1. Climatic element codes and description.

Code	Description
10	number of days with daily snowfall amount equal to the specified threshold
11	number of days with daily snowfall amount greater than or equal to the specified threshold
20	monthly or seasonal total snowfall amount
40	daily snowfall amount (all days examined, whether they had snowfall or not)
41	daily snowfall amount (only days having snowfall examined)
50	number of days with snow depth equal to the specified threshold
51	number of days with snow depth greater than or equal to the specified threshold
80	daily snow depth amount (all days examined, whether they had snow cover or not)
81	daily snow depth amount (only days with snow cover examined)

Table 2. Statistic codes and description.

Code	Description
MD	median (50th percentile)
MN	average (mean)
MX	maximum (same as G1)
NY	number of years with non-missing data
Q1	first quartile (25th percentile)
Q3	third quartile (75th percentile)

Table 3. Date indicator codes and description.

Code	Description
blank	blank indicates the value is a data value
D	value is a date (day of the month)
Y	value is a year

Table 4. Threshold codes and description.

Code	Description
####	a snowfall or snow depth threshold value in inches and tenths
-8.8	a trace value (a value > 0.0 inches but < 0.05 inches for snowfall, or a value > 0.0 inches but < 0.5 inches for snow depth)
-9.9	threshold does not apply to this record